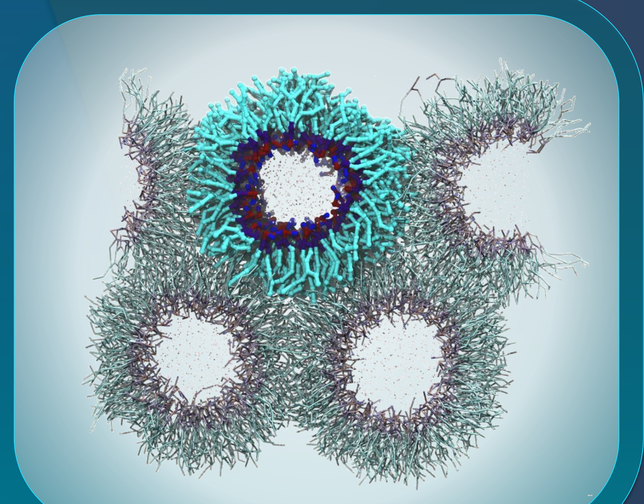
Computational predictions of amphiphile aggregation for early compartmentalisation.

Richard J Gillams[1], Markus Meringer[2] & H James Cleaves II[1,3]

1. Earth-Life Sciences Institute, Tokyo Institute of Technology.
2. Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen.
3. Institute for Advanced Studies, Princeton.

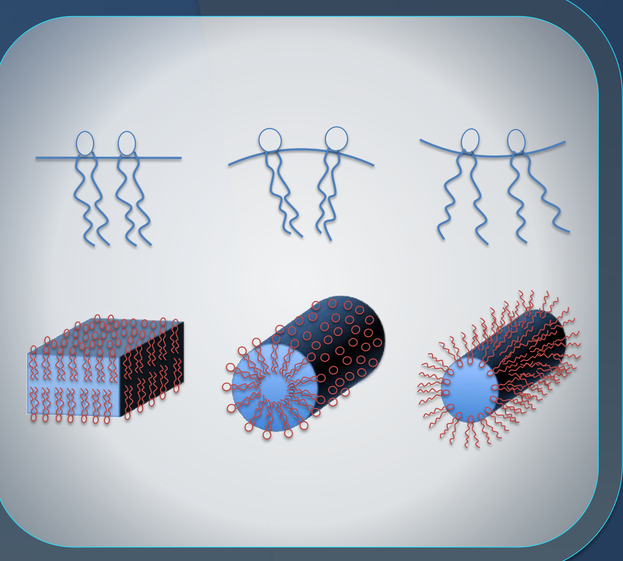
Introduction

- Extant biology uses a vast array of lipids to perform a range of tasks.
- Compartmentalisation is critical for the existence of extant life, providing:
  - Separation of chemical environments.
  - Enhanced local concentration of molecules.
  - Interfaces with reduced dimensionality.
  - The existence of individuals, leading to competition and evolution.
- Want to be able to predict which kind of molecules are able to aggregate to form environments that can harbour/encourage complex chemistry/life.
- Using high resolution models is not compatible with the long length and time scales required to predict self-assembly/aggregation behavior.
- Therefore it is important to develop a computationally efficient way to predict aggregation and screen large compound libraries.



Approach

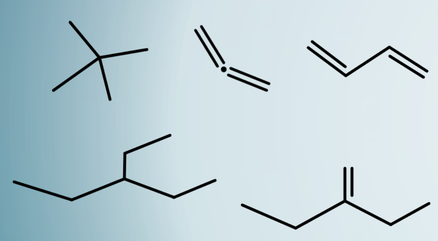
- There are a range of methods available for producing or accessing libraries of molecules.
- Through the recent explosion of lipidomics, there are a number of tools for mass spectrometry which include large compound libraries.
  - LipidBlast.
  - LipidHome.
  - Metabolite databases.
- These give access to biologically relevant lipids, but do not facilitate the identification of novel molecules.
- The work reported here involves the development of an exhaustive library of molecules based on design rules outlined below.
- This library was interrogated in order to identify molecules which possess properties which are commensurate with an ability to form membranes.
- These properties relate to both the propensity for aggregation and the shapes of the aggregates which would be formed.

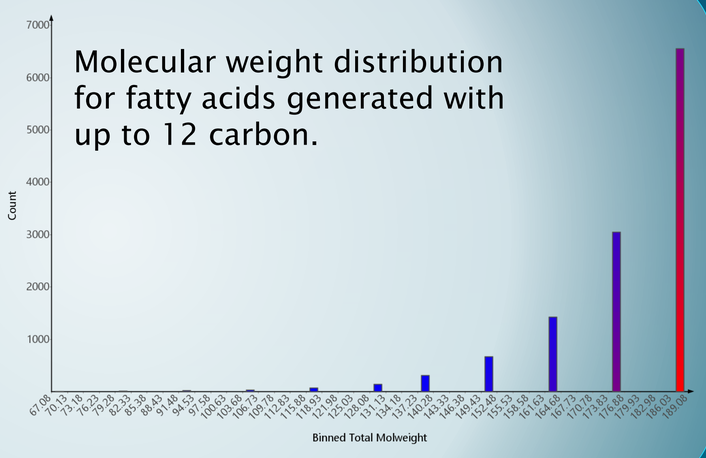


Library generation

- Exhaustive library generation can be achieved with Molgen.
- Within Molgen a set of disallowed molecular motifs and ranges of molecule parameters govern the creation of chemical structures.
- Through this approach, a library of fatty acids can be developed.

Disallowed motifs

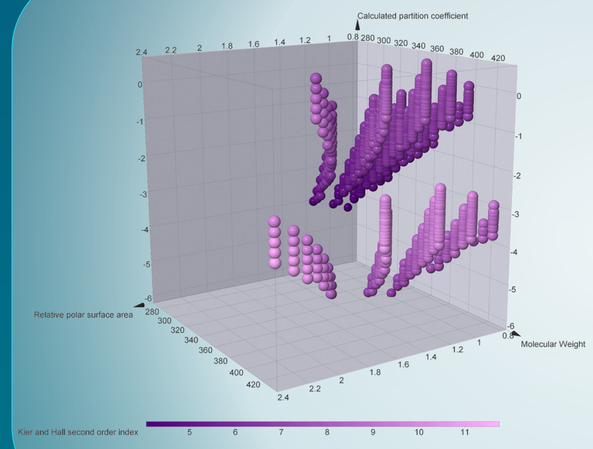




- Due to the amphiphilic nature of self-assembling molecules, the distribution of heteroatoms within the molecules in question is not uniform. This is challenging to overcome in an efficient fashion using Molgen.
- To enhance the efficiency, a computational pipeline has been developed. This involves generating sections of the molecules, which are then combined using a reaction simulation protocol. ChemAxon Reactor is used for this purpose.

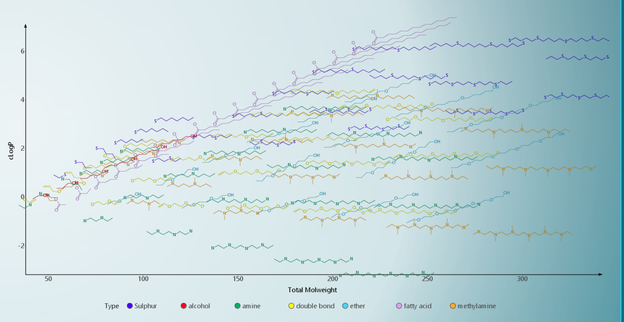
Library interrogation

- Once a library of amphiphilic molecules has been constructed, chemoinformatics approaches can be used to evaluate these structures.
- Chemical descriptors are computationally efficient to calculate and describe the properties of molecules based on chemical structure alone.
- Examples in the scientific literature identify chemical descriptors that have proved effective indicators for aggregation behavior.
- The most successful descriptors have included properties such as the Kier-Hall index, estimated partition coefficients and measures of the polarity of the molecule, as shown below.
- By plotting these descriptors it is possible to assess the coverage of chemical space achieved by the library of molecules.



Lipid library descriptors

Exploring descriptor space by design



Further Work

- Identify ways to gain comprehensive chemical space coverage without relying on exhaustive structure generation.
- Develop QSPR models to predict behaviour.
- Design novel self-assembling systems.
- Test with experiment.

World Premier International Research Center Initiative

東京工業大学 Tokyo Institute of Technology